

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 26, 2001, 10:36:00 ; Search time 13.34 Seconds  
(without alignments)  
1286.292 Million cell updates/sec

Title: US-09-497-967-7  
Perfect score: 2540  
Sequence: 1 MKNILVILISLFINQIKS.....QCDFANFLSILLISYLL 468

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	236.5	9.3	667	1	TS1L_GIALA
2	228.5	9.0	713	1	TSAN_GIALA
3	189	7.4	328	1	C170_GIALA
4	188	7.4	1609	1	LMG1_HUMAN
5	188	7.4	5376	1	ZAN_MOUSE
6	185.5	7.3	1696	1	PKC5_MOUSE
7	184.5	7.3	687	1	VS41_GIALA
8	183.5	7.2	1680	1	FUR2_DROME
9	179.5	7.1	1877	1	PKC5_MOUSE
10	179	7.0	2715	1	GL56_PARPR
11	178.5	7.0	1607	1	LMG1_MOUSE
12	175.5	6.9	3075	1	LMG1_HUMAN
13	174.5	6.9	3635	1	LMA5_MOUSE
14	173	6.8	3712	1	LMA_MOUSE
15	171	6.7	1895	1	YLK3_CAEEL
16	167.5	6.6	2704	1	GL68_PARPR
17	166.5	6.6	969	1	PAC4_MOUSE
18	166	6.5	3084	1	LMG1_MOUSE
19	165	6.5	1700	1	BAR3_CHITE
20	158	6.2	3106	1	LMA2_MOUSE
21	157	6.2	3110	1	LMA2_HUMAN
22	155.5	6.1	1246	1	YMW2_CAEEL
23	155.5	6.1	3672	1	LML2_CAEEL
24	153.5	6.0	1790	1	LMG1_DROME
25	150.5	5.9	670	1	VG50_HSV1
26	150.5	5.9	1169	1	YK82_YEAST
27	150	5.9	1168	1	LMB3_MOUSE
28	148.5	5.8	1639	1	LMG1_DROME
29	147	5.8	712	1	FBL1_CAEEL
30	146.5	5.8	937	1	PAC4_RAT
31	146.5	5.8	1193	1	LMG2_HUMAN
32	145.5	5.7	600	1	SP96_DICTDI
33	145.5	5.7	790	1	ANP_NOTCHO

34	142	5.6	1167	1	XMRK_XIPMA
35	141.5	5.6	610	1	LEM2_HUMAN
36	141.5	5.6	1786	1	LMB1_MOUSE
37	141	5.6	2911	1	FBN2_HUMAN
38	140.5	5.5	1416	1	YH81_CAEEL
39	140	5.5	1064	1	FBP1_STRPD
40	139.5	5.5	677	1	SP87_DICTDI
41	139	5.5	768	1	LEM3_MOUSE
42	138	5.4	2437	1	NOTC_BRARE
43	138	5.4	2871	1	FBN1_MOUSE
44	136.5	5.4	1172	1	LMB3_HUMAN
45	136	5.4	2907	1	FBN2_MOUSE

ALIGNMENTS

RESULT 1					
TS1L_GIALA					
ID	TS1L_GIALA	STANDARD:	PRT:	667 AA.	
AC	Q03185;				
DT	01-JUN-1994 (Rel. 29, Created)				
DT	01-JUN-1994 (Rel. 29, Last sequence update)				
DT	15-JUL-1998 (Rel. 36, Last annotation update)				
DE	MAJOR SURFACE TROPHOZOITE ANTIGEN 11 PRECURSOR.				
GN	TSPI1.				
OS	Giardia lamblia (Giardia intestinalis).				
OC	Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.				
OX	NCBI_TaxID=5741;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ISOLATE AD-1;				
RX	MEDLINE=93241215; PubMed=8479449;				
RA	Ey P.L., Khanna K., Manning P.A., Mayrhofer G.;				
RT	"A gene encoding a 69-kilodalton major surface protein of Giardia intestinalis trophozoites";				
RL	Mol. Biochem. Parasitol. 58:247-258(1993).				
CC	-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE PLASMA MEMBRANE.				
CC	-!- DOMAIN: CONTAINS 27 REPEATS OF A CXXC MOTIF.				
CC	-!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL: M95814; AAA02687.1; -.				
DR	PIR: A48579; A48579.				
DR	HSSP: P02468; IKLO.				
DR	InterPro: IPR000561; EGF-like.				
DR	InterPro: IPR002174; Furin-like.				
DR	SMART: SM00181; EGF 3.				
DR	SMART: SM00261; FU: 5.				
KW	Antigen; Repeat; Transmembrane; Signal.				
FT	SIGNAL 1				
FT	CHAIN 18 667				
FT	DOMAIN 18 633				
FT	TRANSMEM 634 662				
FT	DOMAIN 663 667				
FT	CARBOHYD 591 591				
FT	CARBOHYD 630 630				
FT	SEQUENCE 667 AA; 68475 MW; 1DD9572703232B8D CRC64;				
SQ					

Query Match 9.3%; Score 236.5; DB 1; Length 667;  
Best Local Similarity 22.7%; Pred. No. 6e-10;  
Matches 128; Conservative 49; Mismatches 193; Indels 195; Gaps 30;  
Oy 7 VILIISLFINQIKSANGPVGTETNTAGQVDDLTGTPANCNCQ-----KNFY 52

```

Db 1 MLLAIFYVISTLAKTQTQTCBAECMEVGEITEITRCQTGKVPIDGKCVATANAN 60
QY 53 YNNA-----AFVPGASTCTP-----CP-QKDGAGQPNPAPATANLVQCNV 93
Db 61 CKNASGDGANQVCKMSSVPNTLCTVSPDGVCSVAANEYFVPNADATHDSVVSCE 120
QY 94 KCPAGPA-----TAGGAT-----DYAALITECVNCRINYNENAPNFNAGASTC 137
Db 121 ETPHLLANNKQYGVAGCATCSAPKAPGSDNTPKAATCTKCAAGFL--HTP--SEGLSSC 176
QY 138 -TACPNRVGGALTAGNAATIAQCNVACPTGTALDDGVTTDYVRSFTECVKRLNFYFN 196
Db 177 EETCPGEGYFHTATAESKKT-----CK-STGSS-----EAPNVKIGDCLKC---MYN 222
QY 197 GNGNTPFPNPGKSOCTPCPAIKPANVAQATLGNDA--ITAQCNVA-----CPDG----- 244
Db 223 EASGNT-----LTCEKSAQKPSL-DKTSNCNCTGQNCATFCSSSGGDCGCDSGFILD 275
QY 245 -----TISAAGVNNVAQNTCTNCAFNFNAPNPNPNCSTCLPCPANKDYGA 294
Db 276 GQNCVKSDCXTENCKACTNPKAANEVCTECI-----STHHLTPTSQCVQYCOALGNYYA 329
QY 295 EATAGGAATLAKOCNIA-----CPDGTAIASGATNYVILQTECLNCAANFYFDGNNFQAGS 350
Db 330 GTNADNKKK-CKECTVANCKTND-----QGQCOTCNDGFKYKNGDAGSPCH 374
QY 351 SRCKACPAKVOGAVATAGTATILIAQCALECPAGIVL---TDGTSTYKQ----- 398
Db 375 ESKTCSA-----GTA-----SDCT-ECPTGKALYKNGDGTGTCGCGTTCQGGSG 419
QY 399 -----AASECVK-----AANFYTTKQTDWVAGIDTCTSC----- 428
Db 420 AKTCTGLTIDGASVCSBDQNEYPQNGICTSTARIATVATCKNSNVANGI--CSSOTNGF 477
QY 429 -----NKKLTSGAEAN 439
Db 478 LRMNGGCVETTKPKPGKSVCEGANAD 502

RESULT 2
TS44_GIALA STANDARD; PRT; 713 AA.
ID TSA4_GIALA STANDARD; PRT; 713 AA.
AC P21849;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR SURFACE-LABELED TROPHOZOITE ANTIGEN 417 PRECURSOR.
GN TSA 417.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 30957 / WB;
RX MEDLINE=90280395; PubMed=2352929;
RA Gillin F.D., Hagblom P., Harwood J., Alely S.B., Reiner D.S.,
RT McCafferty M., So M., Guiney D.G.;
RT "Isolation and expression of the gene for a major surface protein of
RT Giardia lamblia."
RL Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).
RN [2]
RP SEQUENCE OF 480-620 FROM N.A.
RC STRAIN=AD-1;
RX MEDLINE=93314970; PubMed=8325510;
RA Ey P.L., Mayrhofer G.;
RT "Two genes encoding homologous 70-kDa surface proteins are present
RT within individual trophozoites of the binucleate protozoan parasite
RT Giardia intestinalis."
RL Gene 129:257-262(1993).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -!- DOMAIN: CONTAINS 29 REPEATS OF THE CXXC MOTIF.

```

```

CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M33641; AAA02688.1; -
DR EMBL; M97488; AAA02581.1; -
DR PIR; A35502; A35502.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00001; EGF; like; 1.
DR SMART; SM00261; FU_3
DR Signal; Antigen; Glycoprotein; Transmembrane; Repeat.
KW SIGNAL
FT CHAIN 1 17
FT CHAIN 18 713
FT DOMAIN 18 679
FT TRANSMEM 680 708
FT DOMAIN 709 713
FT CARBOHYD 289 289
FT CARBOHYD 676 676
FT VARIANT 582 582
FT VARIANT 606 606
FT SEQUENCE 713 AA; 72510 MW; 9AD7195843DE5601 CRC64;
SQ
Query Match 9.0%; Score 228.5; DB 1; Length 713;
Best Local Similarity 24.4%; Pred. No. 2.4e-09;
Matches 122; Conservative 34; Mismatches 170; Indels 175; Gaps 33;
QY 23 CPVGTETNTAGOVDDLTGPANCVQCN-FYFN-----NAAAFVPGASTCTP-----CPQ 71
Db 69 CP-----QHSAGK-----CTCGGNSFMFKDGYSSGEGLPGLSHLCLSSDGDGVCTE 115
QY 72 KKDAGAPN---PPATANLVQCNVK---PAGTATAGGATDYAAIITECVNCRINFYFN 126
Db 116 -----AAPGYFAPVGAAN-TEQSVIACGDTTGTITAAAGNTYKGIADCAEC-----S 161
QY 127 APNFNAGA-----STCTACPNRVGGALTAGNAATIVAQCN----- 162
Db 162 APDATAGAEAGKAVATCTKGVSK-----YLDNVVDVKAQCNSTNKFVAVDDSENGKC 217
QY 163 VACPT---GTALDDGVTTDYVRSFTECVKRLNFYNNNGNTP-----FNP--- 206
Db 218 VSCSDNLNGVANDCTCSYDEQSKIKCTKTDNNYLTSTSEGTSCVQKQCKDGFPPKD 277
QY 207 ---GKSQCTPCPAKP--ANVAQATL--GNDATITAOCHVACPDG----- 244
Db 278 DSSAGNKCLPCNDSTDGIANCATCALVSGRGAALVTC--ACTDGYKPSADKTTCTEAVSN 336
QY 245 --TISAAGVNNVAQNTCTNCAFNFNAPNPNPNCSTCLPCPA--NKDYCAEATAG 300
Db 337 CKTPGCKACSKNEGKNEVCVDCDGSY-----LTP-TSQCIDSCAKIGNYYGA---TEG 386
QY 301 AATLAKOCNIA---CPDGTAIASGATNYVILQTECLNCAANFYFDGNNFQAGSSCKAC 356
Db 387 AKKLCRECTAANKCTCDD-----QGQCACNDGFKYKNGDAGSPCHESCKTC 432
QY 357 PANKVOGAVATAGTATILIAQCALECPAGIVL---TDGTSTYKQ-----AASECVKCA 407
Db 433 SA-----GTA-----SDCT-ECPTGKALYKNGDGTGTCGCGTTCGAGACKTCG 477
QY 408 ANFYTTKQTDWVAGIDTCTSC 428
Db 478 LT-----IDGASYCSEC 489

```

```
RESULT 3
C170_GIALA STANDARD; PRT: 328 AA.
AC FL5799;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SURFACE ANTIGEN CP170 (FRAGMENT).
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 30957 / WB;
RX MEDLINE=88089405; PubMed=3335828;
RA Adam R.D., Aggarwal A., Lal A.A., de la Cruz V.F., McCutchan T.,
RA Nash T.E.;
RT "Antigenic variation of a cysteine-rich protein in Giardia lamblia.";
RL J. Exp. Med. 167:109-118(1988).
CC -!- MISCELLANEOUS: CYSTEINE-RICH, ANTIGENICALLY VARIANT SURFACE
CC PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06741; CAA229916.1; .
DR PIR; S00530;
DR InterPro; IPR002174; Furin-like.
DR SMART; SM00261; FU; 3.
KW Repeat; Antigen.
FT NON_TER 1
FT REPEAT 38 102
FT REPEAT 103 167
FT NON_TER 328 328
SQ SEQUENCE 328 AA: 33438 MW: 373A697A30EDCA21 CRC64;
-----
Query Match 7.4%; Score 189; DB 1; Length 328;
Best Local Similarity 25.9%; Pred. No. 7.5e-07;
Matches 103; Conservative 36; Mismatches 136; Indels 122; Gaps 28;
QY 36 DDLGTPANVCNCKNF--YNNNAAFVPGASTCTPCPKKDGAGQPNPP-----ATANLVT 89
DB 5 DPTGTCVSAVDCQGSAGYTTDS--VSDAKECKK-----NAPCTACAGTADKCT 52
QY 90 QC-----NVKCPAGTAIA-----GGA-----TDYAAITTEVCNCRINFYNENAP 128
DB 53 KCDANGAAPYLKKNTPSDPTGTCVSAVDCQGSAGYTTDSVSDAKECKK-----NAP 105
QY 129 -NFNAG-ASTCTACPVNRVGGALTAGNAATVIAQCNVACPTGT---ALD-----DGVTTD 178
DB 106 CTACAGTADKCTKCDAN-----GAAPYLKKNTPSDPTGTCVSAVDCQGSAGYTTDD 156
QY 179 YVRSTECVCKRLNFYNGNNGNTPPNPGKSQCTPCPAIKPANVAQATLGNDAITIAQCN 238
DB 157 SVSDAKECKKA-----EGQKP-NTAGTQCFSC---SDANCERCD-QND--VCARCS 201
QY 239 VACP--DGTISAAGVNNVNAQNTCTNCAPNFYNNNAPNPNPGNSTCLPCPANKDYGAEA 296
DB 202 TCAPPENKCPAA---TPGCHSCDCTENAMTNQA-----DKCTGCKEGRYLKPES 250
QY 297 TAG--GAATLAKQCNACPDGTAATASCATNWKVILQTECLNCAANFYFDGNFQAGSSRCK 354
DB 251 ANGOSGACLTAECTSDKTHFTREKAGDSKGM-----CLSCS-----DATH---GITGCK 297
QY 355 ACPANKVQGVATAGGTATLIAQALEPC-----PAG 385
DB 298 KC-----ALKTLISGEAESTVVCSS-ECTDKRLTPSG 326
```

```
RESULT 4
LMGL_HUMAN STANDARD; PRT: 1609 AA.
AC FL1047;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LAMININ GAMMA-1 CHAIN PRECURSOR (LAMININ B2 CHAIN).
GN LAMC1 OR LAMB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91093128; PubMed=1985895;
RA Kallunki T., Ikonen J., Chow L.T., Kallunki P., Tryggvason K.;
RT "Structure of the human laminin B2 chain gene reveals extensive
RT divergence from the laminin B1 chain gene.";
RL J. Biol. Chem. 266:221-228(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88198245; PubMed=3360804;
RA Pikkariainen T., Kallunki T., Tryggvason K.;
RT "Human laminin B2 chain. Comparison of the complete amino acid
RT sequence with the B1 chain reveals variability in sequence homology
RT between different structural domains.";
RL J. Biol. Chem. 263:6751-6758(1988).
RN [3]
RP SEQUENCE OF 1393-1609 FROM N.A.
RX MEDLINE=89169663; PubMed=3234037;
RA Fukushima Y., Pikkariainen T., Kallunki T., Eddy R.L., Byers M.G.,
RA Haley L.L., Henry W.M., Tryggvason K., Shows T.B.;
RT "Isolation of a human laminin B2 (LAMB2) cDNA clone and assignment of
RT the gene to chromosome region 1q25-->q31.";
RL Cytogenet. Cell Genet. 48:137-141(1988).
RN [4]
RP SEQUENCE OF 1282-1609 FROM N.A.
RC TISSUE=Endothelial cells;
RX MEDLINE=92216129; PubMed=1806043;
RA Santos C.L.S., Sabbaga J., Brentani R.;
RT "Differences in human laminin B2 sequences.";
RL DNA Seq. 1:275-277(1991).
CC -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ).
CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),
CC LAMININ-6 (K-LAMININ), AND LAMININ-7 (KS-LAMININ).
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

FT	DOMAIN	1038	1609	COILED COIL (POTENTIAL).
FT	DISULFID	286	295	BY SIMILARITY.
FT	DISULFID	288	305	BY SIMILARITY.
FT	DISULFID	307	316	BY SIMILARITY.
FT	DISULFID	319	339	BY SIMILARITY.
FT	DISULFID	342	351	BY SIMILARITY.
FT	DISULFID	344	367	BY SIMILARITY.
FT	DISULFID	370	379	BY SIMILARITY.
FT	DISULFID	382	395	BY SIMILARITY.
FT	DISULFID	398	410	BY SIMILARITY.
FT	DISULFID	400	416	BY SIMILARITY.
FT	DISULFID	418	427	BY SIMILARITY.
FT	DISULFID	430	442	BY SIMILARITY.
FT	DISULFID	445	456	BY SIMILARITY.
FT	DISULFID	447	463	BY SIMILARITY.
FT	DISULFID	465	474	BY SIMILARITY.
FT	DISULFID	477	492	BY SIMILARITY.
FT	DISULFID	724	733	BY SIMILARITY.
FT	DISULFID	726	740	BY SIMILARITY.
FT	DISULFID	742	751	BY SIMILARITY.
FT	DISULFID	754	770	BY SIMILARITY.
FT	DISULFID	773	781	BY SIMILARITY.
FT	DISULFID	775	792	BY SIMILARITY.
FT	DISULFID	795	804	BY SIMILARITY.
FT	DISULFID	807	825	BY SIMILARITY.
FT	DISULFID	828	842	BY SIMILARITY.
FT	DISULFID	830	849	BY SIMILARITY.
FT	DISULFID	852	861	BY SIMILARITY.
FT	DISULFID	864	881	BY SIMILARITY.
FT	DISULFID	884	898	BY SIMILARITY.
FT	DISULFID	886	905	BY SIMILARITY.
FT	DISULFID	907	916	BY SIMILARITY.
FT	DISULFID	919	932	BY SIMILARITY.
FT	DISULFID	935	947	BY SIMILARITY.
FT	DISULFID	937	954	BY SIMILARITY.
FT	DISULFID	956	965	BY SIMILARITY.
FT	DISULFID	968	980	BY SIMILARITY.
FT	DISULFID	983	995	BY SIMILARITY.
FT	DISULFID	985	1001	BY SIMILARITY.
FT	DISULFID	1003	1012	BY SIMILARITY.
FT	DISULFID	1015	1028	BY SIMILARITY.
FT	DISULFID	1031	1031	INTERCHAIN (PROBABLE).
FT	DISULFID	1034	1034	INTERCHAIN (PROBABLE).
FT	DISULFID	1600	1600	INTERCHAIN (PROBABLE).
FT	CARBOHYD	60	60	N-LINKED (GLCNAC. .)
FT	CARBOHYD	134	134	N-LINKED (GLCNAC. .)
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. .)
FT	CARBOHYD	650	650	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1022	1022	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1107	1107	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1151	1161	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1175	1175	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1205	1205	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1223	1223	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1241	1241	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1380	1380	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1395	1395	N-LINKED (GLCNAC. .)
FT	CARBOHYD	1439	1439	N-LINKED (GLCNAC. .)

Query Match 7.4%; Score 188; DB 1; Length 1609;

Best Local Similarity 23.9%; Pred. No. 4.3e-06;

Matches 99; Conservative 28; Mismatches 165; Indels 122; Gaps

Qy	75	AGAPNPPTANILVTCNVKCPAGTAAGGATDYAAII	TECVNCRINFPYNAPNFNAGA	134
Db	674	ASARPGGVPAWVWESCT--CPVG---YGGQF-----	CEMC-LSGYRRETNLGP-Y	718
Qy	135	STTCTACPVNRVGGALTAGNAATIVAQOC--	NVACPTGTALDDGVITDYVRSFTECVKCRLN	192
Db	719	SPCVLCACN---GHSETCDPETGVNCRCRNTFAGP-	-----HCEKSDG	757
Qy	193	FYNGNNGNTFPNPGSKQCTPCP-----AIKP-----	ANVAQATFLGNDAITTAQCN	238



FT	CARBOHYD	1908	1908	N-LINKED (GLCNAC. . .)	(POTENTIAL)	AC
FT	CARBOHYD	1933	1933	N-LINKED (GLCNAC. . .)	(POTENTIAL)	DT
FT	CARBOHYD	2028	2028	N-LINKED (GLCNAC. . .)	(POTENTIAL)	DT
FT	CARBOHYD	2111	2111	N-LINKED (GLCNAC. . .)	(POTENTIAL)	DE
FT	CARBOHYD	2142	2142	N-LINKED (GLCNAC. . .)	(POTENTIAL)	DE
FT	CARBOHYD	2332	2332	N-LINKED (GLCNAC. . .)	(POTENTIAL)	GN
FT	CARBOHYD	2533	2533	N-LINKED (GLCNAC. . .)	(POTENTIAL)	OS
FT	CARBOHYD	2575	2575	N-LINKED (GLCNAC. . .)	(POTENTIAL)	OC
FT	CARBOHYD	2692	2692	N-LINKED (GLCNAC. . .)	(POTENTIAL)	OC
FT	CARBOHYD	2812	2812	N-LINKED (GLCNAC. . .)	(POTENTIAL)	OX
FT	CARBOHYD	3052	3052	N-LINKED (GLCNAC. . .)	(POTENTIAL)	RP
FT	CARBOHYD	3065	3065	N-LINKED (GLCNAC. . .)	(POTENTIAL)	RP
FT	CARBOHYD	3144	3144	N-LINKED (GLCNAC. . .)	(POTENTIAL)	RX
FT	CARBOHYD	3172	3172	N-LINKED (GLCNAC. . .)	(POTENTIAL)	RT
FT	CARBOHYD	3288	3288	N-LINKED (GLCNAC. . .)	(POTENTIAL)	RT
FT	CARBOHYD	3292	3292	N-LINKED (GLCNAC. . .)	(POTENTIAL)	RL
FT	CARBOHYD	3782	3782	N-LINKED (GLCNAC. . .)	(POTENTIAL)	CC
FT	CARBOHYD	4005	4005	N-LINKED (GLCNAC. . .)	(POTENTIAL)	CC
FT	CARBOHYD	4136	4136	N-LINKED (GLCNAC. . .)	(POTENTIAL)	CC
FT	CARBOHYD	4243	4243	N-LINKED (GLCNAC. . .)	(POTENTIAL)	CC
FT	CARBOHYD	4254	4254	N-LINKED (GLCNAC. . .)	(POTENTIAL)	CC
FT	CARBOHYD	4335	4335	N-LINKED (GLCNAC. . .)	(POTENTIAL)	CC
FT	CARBOHYD	4376	4376	N-LINKED (GLCNAC. . .)	(POTENTIAL)	CC
FT	CARBOHYD	4586	4586	N-LINKED (GLCNAC. . .)	(POTENTIAL)	CC
FT	CARBOHYD	5136	5136	N-LINKED (GLCNAC. . .)	(POTENTIAL)	CC
FT	CARBOHYD	5252	5252	N-LINKED (GLCNAC. . .)	(POTENTIAL)	CC
SQ	SEQUENCE	5376 AA; 579908 MW; 0E44DB77DF2A2620 CRC64;				
Query Match						
Best Local Similarity 7.4%; Score 188; DB 1; Length 5376;						
Matches 114; Conservative 52; Mismatches 184; Indels 188; Gaps 30;						
QY	17	QIKSANCPTGTETAGQVDDLGTPANCVN-----CQKNFYNNAAAFVPGASTCTPP-C--	69			
Db	3151	QCONFOCLTKYCK-----DLKDGSSNCTNIPLOCPAHSRYTNC-----LPSCPLCLD	3199			
QY	70	POKDKAOPNPATANLVTCQNV-----KCPAGTAIAGGATDYAAIIECVNCRNF	122			
Db	3200	PEGLCGTSPKVPSTCREGCICOPGYLMHKNK-----VLRIFCGCKNTQAGF	3247			
QY	123	YNENAPNFNAGASTCTACPNRVGGGALTAGNATIAVQAC-NVACPTGTALDGVTTDYVR	181			
Db	3248	ISADKTWISRGCTQSCTCP-----AGAI-----HCRNFKCPSTG-----	3281			
QY	182	SFTECVKRLNEYNGNGTTFN-----PGKSQCTPC-----PAIKP	219			
Db	3282	-----YCKNGDSSNCTEITLQCPNTSQTDLCLPSCVPSCSNRCEVTSPV-P	3329			
QY	220	ANVAATLGNDAI-----TAQCNVACPDGTISAAGVNNVVAQ-----NTE	260			
Db	3330	SSCREGLCNHGFVSEDKVPRPTQCGCKDARGAIIIPAG-KTWTSGTQSCACVGNIIQ	3388			
QY	261	CTN--CAPNFYNNAPNFNGNSTC-----LPCPANKDYGAETA-----GGAATLA	305			
Db	3389	CONFOCPPEY-----CKDNSEGSSTCTKITLQCPAHTQYTSCLPCLDPSELCKDIS	3445			
QY	306	KOCNIACPDGTAIASGATNVIIQTECLNCAANFYFDGNNFQAGSSRCKACAPANKVQGV	365			
Db	3446	PKVSTCKEGVCQSG-----YVLNSDKCVLRA-----BCDCKDAQAL	3484			
QY	366	ATAGGTATL-----IAQC-ALECPAGTVLTDGTTSTVYKQASCEVK-----CAA	408			
Db	3485	IPAGKTWTSPTGCTQSCACMGAVQCQSQCPPGTYCKDN-----EDGNSNCAKITLOC	3539			
QY	409	NFYTKQTDWAGIDTC-TSC--NKKLTSGAEANLPSAKNNIQCDFANFLISLLLI	463			
Db	3540	HSLFTN-----CLPPCLPCLDPPDLGCKGASPKVPSTCKEGCICQSGYVLSNNKCLL	3591			
RESULT 6						
PCK5_BRACL						
ID	STANDARD; PRT; 1696 AA.					

AC	Q9NJ15; Q9NJ16; Q9NJ14;
DR	20-AUG-2001 (Rel. 40, Created)
DT	20-AUG-2001 (Rel. 40, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	PROTEIN CONVERTASE SUBTILISIN/KEXIN TYPE 5 PRECURSOR (EC 3.4.21.-)
DE	(PROTEIN CONVERTASE PC6-LIKE) (APC6).
GN	PC6.
GN	Branchiostoma californiensis (California lancelet) (Amphioxus).
OC	Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC	Branchiostoma.
OX	NCBI_TaxID=7738;
RP	[1]
RP	SEQUENCE FROM N.A. (ISOFORM A; ISOFORM B AND ISOFORM C).
RA	MEDLINE=20175281; PubMed=10708868;
RA	Oliva A.A. Jr., Chan S.J., Steiner D.F.;
RT	"Evolution of the prohormone convertases; identification of a
RT	homologue of PC6 in the protochordate amphioxus.";
RL	Biochim. Biophys. Acta 1477:338-348(2000).
CC	-!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
CC	WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
CC	OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (3Y SIMILARITY).
CC	-!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
CC	PROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-1-ZAA BONDS, WHERE XAA
CC	CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
CC	-!- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.
CC	ISOFORM B IS A TYPE I MEMBRANE PROTEIN.
CC	-!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A, B (SHOWN HERE) AND C; ARE
CC	PRODUCED BY ALTERNATIVE SPLICING.
CC	-!- DOMAIN: THE PROPEPTIDASE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
CC	ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
CC	RETICULUM.
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC	SUBTILASE FAMILY.
CC	-!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; AF184615; AAF26300.1; -
DR	EMBL; AF184616; AAF26301.1; -
DR	EMBL; AF184617; AAF26302.1; -
DR	InterPro: IPR000561; EGF-like.
DR	InterPro: IPR002174; Furin-like.
DR	InterPro: IPR002884; P_domain.
DR	InterPro: IPR000209; Peptidase_S8.
DR	ProDom: PD000717; P_domain; 1.
DR	SMART: SM00181; EGF; 2.
DR	SMART: SM00261; FU; 17.
DR	PROSITE; PS00136; SUBTILASE ASP; FALSE_NEG.
DR	PROSITE; PS00137; SUBTILASE HIS; 1.
DR	PROSITE; PS00138; SUBTILASE_SER; 1.
KW	Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW	Cleavage on pair of basic residues; Repeat; Alternative splicing;
KW	Transmembrane.
FT	SIGNAL
FT	PROPEP
FT	CHAIN
FT	1 25
FT	26 110
FT	111 1696
FT	POTENTIAL.
FT	POTENTIAL.
FT	PROTEIN CONVERTASE SUBTILISIN/KEXIN
FT	TYPE 5.
FT	EXTRACELLULAR (POTENTIAL).
FT	POTENTIAL.
FT	CYTOPLASMIC (POTENTIAL).
FT	CATALYTIC.
FT	HOMO B.
FT	CYS-RICH MOTIF (CRM) REGION.
FT	CLEAVAGE (AUTO-) (BY SIMILARITY).
FT	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CHARGE RELAY SYSTEM (BY SIMILARITY).

```
FT CARBOHYD 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 885 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1259 DDITLRCGICITSCGPGGYMDRREKCKACHPTCKECSDEY
FT DDITACNDGFLTLTDASSCEAGCP -> AENQNOASFQCPFA
FT PREVSVLAEALGHLRLYSLTDVPPQSNPPDVTVLGADRRL
FT TTATSAAGRA (IN ISOFORM C).
FT VARSPLIC 1324 CHPTCKECSDEYDDTCTACNDGFLTLTDASSCEAGCPGQFL
FT VARSPLIC 1288 HUGDCDSCHREKTC -> IARCVDRRDRSKDVLVRFNFC
FT VRYIFVKRCGCKLYLMEDRPMRGRSSQPTOGRN (IN
FT ISOFORM A).
FT VARSPLIC 1344 MISSING (IN ISOFORM A).
FT SEQUENCE 1696 AA; 188410 MW; 281CBEL784257CBD CRC64;

Query Match 7.3%; Score 185.5; DB 1; Length 1696;
Best Local Similarity 20.4%; Pred. No. 6.9e-06;
Matches 101; Conservative 43; Mismatches 189; Indels 161; Gaps 25;

QY 18 IKSANCPVGTETTAG-----QVDDLGTPANVCNCKNFYNNAAAFVPGASTCT 67
DB 1048 VRTNCPSFTYDDQRECRPHDNEACDGPNNQNCSCKEGYKT-----PDGCS--T 1100
QY 68 PCFQ---KIDAGQPNP-----PATANLV-----TQCNVCKPAGTAIAGG 104
DB 1101 GCPNRYKDDTNKECKPCDSCFTCSGPASFCLSCADGDFLHESCRSTCPAG--FIGN 1158
QY 105 ATDYAALITECVNCRINFYNENAPNFNAGASTCTACPNRVGGALTAGNAATVACQNV 164
DB 1159 AESHECVESSC-----EQQYYSSETGCEDCPYN-----CRACDNDGDCAEC 1201
QY 165 CPTGTALD-----DGVTTDYVRSFTECVKCRLENYNGNNGTPEFNGKSOCTPC 214
DB 1202 APTIYVVGRCRPEETCEDEGYDRDRTAE--LSCR-----PCHOSCKTC 1245
QY 215 PALKPANVAQALGNDAITTAQCNVACPGDTTISAAGVNNVAQNTFC--TNCAPNFY--NNN 272
DB 1246 SG--PSD-----TDCDSCKGDDTILDRG-----ECITSCGPGGYMDRR 1281
QY 273 APNFNPGNSTCLPCPANKDYGAETA---GGAATLAKOCNIACPDGTATASCATNYVILQ 329
DB 1282 EKKCKACHPTCKEC--SDEYDDTCTACNDGFLTLTDASSCEAGCPGQFLHHG----- 1331
QY 330 TECLNCAANFYDGNFQAGSRCKACPAKPVQGAVATAGGTATLIAQALCEPCPAGTVLT 389
DB 1332 -DCDSC-----HRECKTCDGPHDNCILSCOPGYLNDQCSHCPCGTF-- 1374
QY 390 DGTSTYKQAASE---CVKCAANFYTTKQTDWVAGIDTCTSC-----NKKLTSGA 436
DB 1375 ---EETIEDDSGETVLQCRLCHVNCKTCHG-----EGEEDCMECANDIKYKQDGRVCVTECQ 1427
QY 437 EANLPESAKKNQOC 450
DB 1428 EGHYPDLTNECQOC 1441

RESULT 7
VS41_GIALA
ID VS41_GIALA STANDARD; PRT; 687 AA.
AC P92127;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
OS VARIANT-SPECIFIC SURFACE PROTEIN VSP4A1 PRECURSOR (CRISP-90).
OC Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O2-4A1;
RX MEDLINE=97321554; PubMed=9178264;
RA Papanastasiou P., Bruderer T., Li Y., Bommeli C., Koehler P.;
```

```
RT "Primary structure and biochemical properties of a variant-specific
RL surface protein of Giardia."
RN Mol. Biochem. Parasitol. 86:13-27(1997).
RX CHARACTERIZATION.
RX MEDLINE=97233006; PubMed=9078242;
RA Papanastasiou P., McConville M.J., Ralton J., Koehler P.;
RT "The variant-specific surface protein of Giardia, VSP4A1, is a
RL glycosylated and palmitoylated protein."
CC Biochem. J. 322:49-56(1997).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -!- PTM: O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC
CC AT THE REDUCING TERMINUS.
CC -!- PTM: PALMITOYLATED.
CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 283743; CAB06038.1; -
CC GlycoSuiteDB; P92127; -
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002174; Furin-like.
CC SMART; SM00181; EGF; 2.
CC SMART; SM00001; EGF_Like; 1.
CC SMART; SM00261; FU; 3.
CC Antigen; Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate;
KW SIGNAL.
KW CHAIN. 1 14
KW DOMAIN 15 687
KW TRANSMEM 661 681
KW DOMAIN 682 687
KW SEQUENCE 687 AA; 70857 MW; D892F675D626D7EC CRC64;
SQ
Query Match 7.3%; Score 184.5; DB 1; Length 687;
Best Local Similarity 21.1%; Pred. No. 3.3e-06;
Matches 119; Conservative 54; Mismatches 214; Indels 177; Gaps 29;

QY 12 SLFINQIKSANCPVGTET---NTAGQVDDLGTPANVCNCKNFYNNAAAFVPGASTCT 67
DB 124 SOYIFQNKATPSEKSGECSILCWDITDRGVNGV-ANCATC-----TAPASSTGPATCT 175
QY 68 PCP---QKKDAGAQPNPPATANLVTCNVKCPAGTAIAGGATDYAALITECVNCRINFY 123
DB 176 ECMAGTYKKSD-----TEC-AACHSDCATCSGEAN-----NQCTSCETGKY 215
QY 124 -----NENAPFN---AGASTCTACPNRVGGALTAGNAATVACQNVACPTG---- 168
DB 216 LKSNQCVKNTCTNHYPPDDTSMTCVACTVLDANCATCFSDSATKAGKC-LTCNSKNIPR 274
QY 169 TALDGVVTTDYVRSFTECVKCRLENYNGNNGTPEFNGKSOCTPCPAIKPANVAQATLG 228
DB 275 TTL-DGTSTCVENSYAGC-----QGADNELFMKEDQSACLLCGDTKEASNDKG-VA 323
QY 229 NDATITTAQCN-----VACPDGTISAAG-VNNVVAQN-----TECTNCAPNFYNN 271
DB 324 NCRTCTKNANDSPPTCTACLDGYFLERGSCSTTTTCADNCATCSEATTEDCKICKAGFF-- 381
QY 272 NAPNFNPGNSTCLPCPANKDYG---AEATAGGAATL-AKQC-----NIACPDG 315
DB 382 ---LASPEGCKCISCSDTNGGIDGCAECTKEPAGPLKCTKAPNRRKPGAGTSDNYTCTEK 438
QY 316 TAI-----ASGATNYVILQTE-----C 332
DB 439 TCEDPTVCGGTSGACDAIVIDANGKEHYVCSYCGETNKKFPIDGLCTDNKGTNAGCTDHTC 498
```

PROSITE; PS00138; SUBTILASE\_SER; 1.  
Hydrolase; Serine protease; Glycoprotein; Signal; Transmembrane;  
Multigene family; Zymogen.

Multigene family, $\lambda$ phage	Signal	Potential
PROPEP	319	POTENTIAL.
CHAIN	1680	POTENTIAL.
ACT_SITE	418	FURIN-LIKE PROTEASE 2.
ACT_SITE	457	CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE	457	CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE	638	CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE	638	10 X TANDEM REPEATS, CYS-RICH.

DOMAIN	962	1444
REPEAT	1062	1007
1.	962	1007
2.	1008	1057
3.	1058	1104
4.	1105	1153
5.	1154	1205
6.	1206	1254
7.	1255	1299

REPEAT	1200	8.
REPEAT	1300	1346
REPEAT	1347	1393
REPEAT	1394	1444
REPEAT	1444	10.
TRANSMEM	1508	1552
POTENTIAL		POTENTIAL
POTENTIAL		POTENTIAL

TERMINAL	1680	1533	1680	CYTOPLASMIC (POTENTIAL)
DOMAIN	3	3		N-LINKED (GLCNAC. .)
CARBOHYD				N-LINKED (GLCNAC. .)
CARBOHYD	109	109		N-LINKED (GLCNAC. .)
CARBOHYD	130	130		N-LINKED (GLCNAC. .)
CARBOHYD				N-LINKED (GLCNAC. .)

CARBOHYD	203	203	N-LINKED (GLCNAC..)	(POTENTIAL)
CARBOHYD	443	443	N-LINKED (GLCNAC..)	(POTENTIAL)
CARBOHYD	481	481	N-LINKED (GLCNAC..)	(POTENTIAL)
CARBOHYD	928	928	N-LINKED (GLCNAC..)	(POTENTIAL)
CARBOHYD	1063	1063	N-LINKED (GLCNAC..)	(POTENTIAL)

CARBOHYD	1061	N-LINKED (GLCNAC..)	(POTENTIAL)
CARBOHYD	1061	N-LINKED (GLCNAC..)	(POTENTIAL)
CARBOHYD	1182	N-LINKED (GLCNAC..)	(POTENTIAL)
CARBOHYD	1275	N-LINKED (GLCNAC..)	(POTENTIAL)
CARBOHYD	1278	N-LINKED (GLCNAC..)	(POTENTIAL)
CARBOHYD	1278	N-LINKED (GLCNAC..)	(POTENTIAL)
CARBOHYD	1440	N-LINKED (GLCNAC..)	(POTENTIAL)
CARBOHYD	1440	N-LINKED (GLCNAC..)	(POTENTIAL)

SEQUENCE	1680 AA;	183599 MW;	0A99CE8770A8E293	CRC64;
CARBOHYD	1440	1440	N-LINKED (GLYCNA...	7 28; Score 183.5; DB 1; Length 1680;

Very Match  
Best Local Similarity 19.3%; pred. No. 9.5e-06;  
Mismatches 152; Indels 213;  
Conservative 99; Conservative 49;

23 CPVGTETNTAGVDDUGLTPANGVYRCQAGKYL  
1051 CPDGFYFNSRNRCTVCPCEP-NCASQDHPEYCTSCDHLVMHEHKYCSACPLDITYETED  
56 AAFAFPGASTCTPCQKKDAGAQPNPPATANLVT-----QCNVKCPAGTAIAGG

56 AAAFVFGASCLIFCPQADAGQ... : : : : :  
1110 KCAFC--HSTCATC-----NGPTDQCITCRSRYAWQNKLCISCPDGF-----  
106 TDYA-AIITECVNCRINFYNNAPNFNAGASTCTA-----CPVNRV-----

106 DIA ALLCNCNANLNNLNKNDKCIIVSGS  
||| |||  
EGCTCTNGVCSECLQNTWLNKRDKCIIVSGS  
1152 --YADKKRLECMPCQ-----  
  
147 GALTAGNAATIVAQN-----VACPTGTAIDGGVTTDYVRSFTECVNYV

1198 GCSEFYSQVEGQCRPHASGSGNGPADTCTCTCPPNRLLEQ-----SRCV  
189 CRNFYYNGNGNTFPNPKSCTPC-----PAIKPANVAQTGLNDATITAQCNVAC

1248 CREGFFVEAG-----SUCSPCLHTCSQCVSTNCNSCKGLELQ-NGECRTTCGAA  
244 GTLSAAGVNNVWQAQNTCTNCAPNFYNNNAPNPNPCGNSTCLPCPANKDYCAEATAGAGAA

1297 GYSDRG-----CAKYLSCHTCSGPRN---QCVCQPA---GWQLAAG---  
304 LAKQCNACPDGTAIASGATNYVILQTECLNCAANEYFDGNPNQAGSSRCKACPANKV

1336 ---ECHPEPEG-----FYKDFGCKC--HHYCKTN-DAGPLACTSCPPHSM  
364 AVATAGGTATLIAQCALECPAGVLTDTGTTSTYK

1379 -----LDGGLMEC-LSSQYDTTATCKTCHDSCRSIFGPGQFSCKGCVPLP

333 INCAANFYFDGNNFOAGSSRCKACPANKYQAVATAGGTALLIA-----QC 378

[illegible]

b

379 ALEC--PAGIVLIDIGLSIRKQSGVGGGAAAR... : |::|  
555 VLACNPLGTI-AGGNAYVGEGSQCTAPDA-----RADGMVAATCTACEDKKPG-606  
  
437 EANLPESA KKNIOCDFANFLISIL #60

437 EANLFESARRNLQDQAFANLQALQ 438  
 : | : | | : :  
 607 -----KSGTGCVCAPDANCKSTM 625

RESULT	8
UR2_DROME	
D_FUR2_DROME	STANDARD;
C_P30432;	PRT; 1680 AA.

01-APR-1993 (Rel. 25, Created)  
01-APR-1993 (Rel. 25, Last sequence update)  
15-APR-1993 (Rel. 25, Last annotation update)  
01-JUL-1998 (Rel. 36, Last annotation update)  
15-JUL-1998 (Rel. 36, Last annotation update)  
FURIN-LIKE PROTEASE 2 PRECURSOR (EC 3.4.21.75) (FURIN 2).

FUR2.  
Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_taxid=7227;  
 [1]  
 SEQUENCE FROM N.A.  
 02381036 PubMed=1512259.  
 RP

XX  
AA  
NA  
RA  
RT

MEDLINE=92381036; PubMed=1512259;  
Roebroek A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dumke U.,  
Rantop M., Gateff E.A.F., Leunissen J.A.M., van de Ven W.J.:  
"Cloning and functional expression of Dfurn2, a subtilisin-like  
proteinase processing enzyme of Prosophila melanogaster with multiple  
functional domains".

repeats of a cysteine motif";  
protein processing enzyme of *Drosophila*  
J. Biol. Chem. 267:17208-17215(1992).  
-1- FUNCTION: FURIN IS LIKELY TO REPRESENT THE UBIQUITOUS ENDOPROTEASOMAL  
ACTIVITY WITHIN CONSTITUTIVE SECRETORY PATHWAYS AND CAPABLE OF

-1- CATALYTIC ACTIVITY: RELEASE OF ARG-XAA-YAA-ARG-1-ZAA BONDS, WHERE XAA PROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG OR ARG OR LYS. RELEASES ALBUMIN, CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE SUBTILASIN FAMILY, FURIN SUBFAMILY.

----- SUBILLARS FAMILY, LONNA CO. -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on the use of this database in non-commercial settings.

use by non-profit institutions as long as its content is not modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	EMBL; M94375; AAA28551.1; -
DR	PIR; A43434; A43434.
DR	HSP; Q99405; 1MPT.

DR FlyBase: FBgn0004598; Fur2.  
DR InterPro: IPR000561; bGF-like.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR002884; P\_domain.  
DR InterPro: IPR002884; P\_domain.

DR InterPro; IPR000209; Peptidase\_S8.  
DR pfam; PF00757; Furin-like; 1.  
DR pfam; PF01483; P; 1.  
DR pfam; PF00082; Peptidase\_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.  
DR ProDom; PD000717; P\_domain; 1.  
DR SMART; SM00181; EGF; 1.  
DR SMART; SM00261; FU; 10.  
DR

```
DR PROSITE: PS00136; SUBTILASE_ASP; 1.  
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
```



QY 398 -QAASECVKCAANFYTTKQTDWVAGIDTCTSCN 429  
 Db 1428 DQLNSQCVSCCQNTLAETSSAA-----CCNCD 1456

RESULT 9  
 PC5\_MOUSE  
 ID PC5\_MOUSE STANDARD; PRT: 1877 AA.  
 AC 004592; Q62040;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROTEIN CONVERTASE SUBTILISIN/KEXIN TYPE 5 PRECURSOR (EC 3.4.21.-)  
 DE (PROTEIN CONVERTASE PC5) (SUBTILISIN/KEXIN-LIKE PROTEASE PC5)  
 DE (CONVERTASE PC5) (PC6) (SUBTILISIN-LIKE PROTEIN CONVERTASE 6)  
 DE (SPC6).  
 GN PCSK5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PC5B).  
 RC STRAIN=ICR; TISSUE=Intestine;  
 RX MEDLINE=93327934; PubMed=8335106;  
 RA Nakagawa T., Murakami K., Nakayama K.;  
 RT "Identification of an isoform with an extremely large Cys-rich region  
 of PC6, a kex2-like processing endoprotease.";  
 RL FEBS Lett. 327:165-171(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM PC5A).  
 RC TISSUE=Brain, and Intestine;  
 RX MEDLINE=933224489; PubMed=8468318;  
 RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K.,  
 RA Nakayama K.;  
 RT "Identification and functional expression of a new member of the  
 mammalian kex2-like processing endoprotease family: its striking  
 structural similarity to PACE4.";  
 RL J. Biochem. 113:132-135(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM PC5A).  
 RC TISSUE=Adrenal cortex;  
 RX MEDLINE=93342056; PubMed=8341687;  
 RA Lusson J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;  
 RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a  
 candidate proprotein convertase expressed in endocrine and  
 nonendocrine cells.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).  
 RN [4]  
 RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.  
 RX MEDLINE=97103178; PubMed=8947550;  
 RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K.,  
 RA Bendayan M., Seidah N.G.;  
 RT "The isoforms of proprotein convertase PC5 are sorted to different  
 subcellular compartments.";  
 RL J. Cell Biol. 135:1261-1275(1996).  
 RN [5]  
 RP DEVELOPMENTAL EXPRESSION.  
 RX MEDLINE=96293359; PubMed=8698813;  
 RA Constam D.B., Calton M., Robertson E.J.;  
 RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone  
 morphogenetic proteins at distinct sites during embryogenesis.";  
 RL J. Cell Biol. 134:181-191(1996).  
 RN [6]  
 RP DEVELOPMENTAL EXPRESSION.  
 RX MEDLINE=97436919; PubMed=9291583;  
 RA Rancourt S.L., Rancourt D.E.;  
 RT "Murine subtilisin-like proteinase SPC6 is expressed during embryonic  
 implantation, somitogenesis, and skeletal formation.";  
 RL Dev. Genet. 21:75-81(1997).  
 CC -1- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY  
 WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE  
 OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE

CC FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED  
 CC IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION  
 CC OF GROWTH FACTORS.  
 CC -1- CATALYTIC ACTIVITY: RELEASE OF NATURE PROTEINS FROM THEIR  
 CC PROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA  
 CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.  
 CC -1- SUBCELLULAR LOCATION: PC5A IS SECRETED THROUGH THE REGULATED  
 CC SECRETORY PATHWAY. PC5B IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO  
 CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH  
 CC EARLY ENDOSOMES.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS: PC5B/LONG (SHOWN HERE)  
 CC AND PC5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: PC5A IS EXPRESSED IN MOST TISSUES BUT IS MOST  
 CC ABUNDANT IN THE INTESTINE AND ADRENALS. PC5B IS EXPRESSED IN THE  
 CC INTESTINE, ADRENALS AND LUNG BUT NOT IN THE BRAIN.  
 CC -1- DEVELOPMENTAL STAGE: WEAKLY EXPRESSED THROUGHOUT THE EMBRYO.  
 CC EXCEPT IN THE DEVELOPING NERVOUS SYSTEM, THE RIBS AND THE LIVER,  
 CC BUT MARKEDLY UPREGULATED AT DISCRETE SITES DURING DEVELOPMENT. AT  
 CC E6.5, PROMINENT EXPRESSION OBSERVED IN DIFFERENTIATED DECIDUA. AT  
 CC E7.5, INTENSE EXPRESSION IN EXTRAEMBRYONIC ENDODERM, AMNION AND  
 CC NASCENT MESODERM. AT 8.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK  
 CC SAC FOLLOWED BY A CONFINATION TO DERMATOTOME COMPARTMENT BETWEEN  
 CC E9.5 AND E11.5. ABUNDANT EXPRESSION IN AER (THICKENED ECTODERMAL  
 CC CELLS OF LIMB BUDS). AT E12.5, EXPRESSION IN THE LIMBS IS CONFINED  
 CC TO THE CONDENSING MESENCHYM SURROUNDING THE CARTILAGE. AT THIS  
 CC STAGE, STRONG EXPRESSION ALSO DETECTED IN VERTEBRAL AND FACIAL  
 CC CARTILAGE PRIMORDIA AND IN THE MUSCLE OF THE TONGUE. AT E16.5,  
 CC ABUNDANT EXPRESSION IN EPITHELIAL CELLS OF THE INTESTINAL VILLI.  
 CC ISOFORM A IS MOST ABUNDANT AT ALL STAGES BUT SIGNIFICANT LEVELS OF  
 CC ISOFORM B OCCUR AT E12.5.  
 CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE  
 CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC  
 CC RETICULUM.  
 CC -1- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN  
 CC SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS  
 CC WITH THE TGN SORTING PROTEIN PACS-1.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE  
 CC SUBTILASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D17583; BAA04507.1; -;  
 CC EMBL: D12619; BAA02143.1; -;  
 CC EMBL: L14932; AAA74636.1; -;  
 CC FIR: JX0248; JX0248.  
 CC PIR: A48225; A48225.  
 CC HSP: Q99405; LMPT.  
 CC MEROPS: S08.076; -;  
 CC MGD: MGI:97515; PCSK5  
 CC InterPro: IPR000561; EGF-like.  
 CC InterPro: IPR002174; Furin-like.  
 CC InterPro: IPR002884; P-Domain.  
 CC InterPro: IPR000209; Peptidase\_S8.  
 CC Pfam: PF01483; P; 1.  
 CC Pfam: PF00082; Peptidase\_S8; 1.  
 CC PRINTS: PR00723; SUBTILISIN.  
 CC PRODOM: PD000717; P-Domain; 1.  
 CC SMART: SM00181; EGF; 3.  
 CC SMART: SM00001; EGF-like; 2.  
 CC SMART: SM00261; FU; 22.  
 CC PROSITE: PS00136; SUBTILASE ASP; 1.  
 CC PROSITE: PS00137; SUBTILASE HIS; 1.  
 CC PROSITE: PS00138; SUBTILASE SER; 1.  
 CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;  
 CC Cleavage on pair of basic residues; Repeat; Alternative splicing;  
 CC Transmembrane.

	Query Match	7.18;	Score 179.5;	DB.1;	Length 1877;	
	Best Local Similarity	20.34;	Pred. No. 2.le-05;			25;
	Matches	96;	Conservative	Mismatches 166;	Indels 163; Gaps	
QY	42	ANCVCNQKNFYNNAAAFVPGASTCTPCPQKDKAGAPNPATANLVITQCNVKCPAGT--	99	:	:	:
DB	1214	SSCKTC-----NGSLCASCTGMYLWLQ-----ACVPSCPQGTPW	1248	:	:	:
QY	100	AATAGATDYAALITECVNCRINFYNENAFPNAGASTCAC---PVNRVGGALTAGNAAT	156	:	:	:
DB	1249	SVTSGSCEKS--EDCVSC-----SGADLCQQCLSPDNIL--LLHSG----	1287	:	:	:
QY	157	IVAOCNVACTPTALDDGVTTDVRSETECVKRLNFYNGNGNTPPNPGKSQCTPCPA	216	:	:	:
DB	1288	---RCVHSCEGFYAKGV-----CEHC-----SSPKTCBGNATSCNS	1323	:	:	:
QY	217	KIPANVAQTALNDATITIAQCNVACPDPGTISAAGYNNWVAQNTE-----CTNCAPNY	269	:	:	:
DB	1324	CE-----GDFVLHDHGVCWKTKPEKHVAVEGVKCHPERCQDCIHEKTKECMPDEF	1374	:	:	:
QY	270	NNN-----APNFPNGNSTCLPCPAN-----KDYGAETAGGAATLAK--OCNIACPD	314	:	:	:
DB	1375	LYNDMCHRSPKSFYFDMRQCVPHKCNECGNKPEDDKCVKADTSKALHNGICLDECPE	1434	:	:	:
QY	315	GT-----AIAGA-----TNVILTQTECL--NCAAFYFDGNNFQA	348	:	:	:
DB	1435	GTYKBEENDCDPCPSCLICSASWTCLACREGFTVVHVDVTAPEKCAAVEIWD-----E	1489	:	:	:
QY	349	GSSRKAC--PANKVOG-----AVATAGGTATLIAQALECPAG-----TVLTDTGT	392	:	:	:
DB	1490	GSHRCQPCHKKSRCSGPSDDCYTCPREFTLNNTCVKECPESGYHTDKDSQQCVLCHSS	1549	:	:	:
QY	393	TSTYKQAAS--ECVKCAANFYTHKQ-----TDWVGIDT--CTSCNKKLTS	434	:	:	:
DB	1550	CRTCEGPHSMOCLSRPGWFOLGKECLLOCROGYIGYSTGSRCREKDKSCKS	1601	:	:	:

```
Query Match          7.0%; Score 179; DB 1; Length 2715;
Best Local Similarity 23.7%; Pred. No. 3.2e-05;
Matches 125; Conservative 38; Mismatches 198; Indels 166; Gaps 32;
```

	QY	20	SANCPVGTETNTAGQVDGLGTCANCVKCNKYNNAA--AFVPGA-----STC----	TP	68
Db	:	: :		:	
	970	AANCFRSSASGTAGYC-----AMTNCO\$-----VTSAECAFAVTGLTGLDHSKCOLYHSS			1020
QY	69	CPOKCD-AQAQ-----PNPPATANLIVTCNVKC-----PAGTAIAG--G			104
Db	1021	CTSLKDGTGCQEYKTTCSGYAATNNCATSGQGCKCFDEVECLRFNSCA\$ITGTGLTAICG			1080
QY	105	ATDYAAII----TEC-----VNCRINFYENNAEFNAGASTCTACPVRNVGGALTAGNAAT			156
Db	1081	TYDAGCVANNVGTFACOEKLATCDL-VLTGNCGSTSAAAADAKC-----AWSGTACL			1131

QY 157 IVAOCNVACP--TGATLDGCVTTDYVRSETECVKCRNLNFFYNGNNGTNPFPKGSQC----- 211  
 Db 1132 AVTTGTHCVYGTGTGLDILCAAY-----NANCTANKAGTACQEKKATCNLY 1179  
 QY 212 ---TPCPAIPKPNVAQATLGNDAITIAQCNVACPDCGTISAAGVNNWV--AQNTCTNCAP 266  
 Db 1180 TTEATCSTSAATAADKCAWSGAACLAIVTVATECAVYVGTGLTDLICAYNANCT----- 1235  
 QY 267 NFYNNAFPNPNSTCLPCPANKD--YGAET--AGGAATLAKQC---NIACPDGTAIA- 319  
 Db 1236 -----ANKAGTACQEKKATCNLYTTEATCSTSAATAADKCAWSGAACLAIVTVAT 1286  
 QY 320 -----SGATNVVILQTECLCAANFFYDGNFQAGSSRCRCACANK----- 360  
 Db 1287 ECAYVGTGLTN-AICAAVYNNCTAN-----RAGT-----ACEKKATCNLYTTEATC 1333  
 QY 361 -VOGAVATA-----GGTATLQAQCALECP--AGTVLTDGTTSTY-----KQAASEC-- 403  
 Db 1334 STSAAATAADKCAWSGAACLAIVTVATECAVYVGTGLTNAICAAVYNNCTANKAGTACQE 1393  
 QY 404 VKCAAFNYTTKOTDWDVAGIDTCTSCNKKLTSCAEANLPESAKKNQOC 450  
 Db 1394 KKATCNLYTTEAT-----CS---TSAATAADKCAWSGAAC 1426

RESULT 11  
 LMGL\_MOUSE  
 ID LMGL\_MOUSE STANDARD; PRT: 1607 AA.  
 AC P02468;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE LAMININ GAMMA-1 CHAIN PRECURSOR (LAMININ B2 CHAIN).  
 GN LMCL OR LAMC-1 OR LAMB-2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88059118; PubMed=3680290;  
 RA Sasaki M., Yamada Y.;  
 RT "The laminin B2 chain has a multidomain structure homologous to the  
 RT B1 chain.";  
 RL J. Biol. Chem. 262:17111-17117(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89000737; PubMed=3167041;  
 RA Durkin M.E., Bartos B.B., Liu S.-H., Phillips S.L., Chung A.E.;  
 RT "Primary structure of the mouse laminin B2 chain and comparison with  
 RT laminin B1.";  
 RL Biochemistry 27:5198-5204(1988).  
 RN [3]  
 RP SEQUENCE OF 1-239 FROM N.A.  
 RX MEDLINE=88228071; PubMed=2836421;  
 RA Ogawa K., Burbello P.D., Sasaki M., Yamada Y.;  
 RT "The laminin B2 chain promoter contains unique repeat sequences and  
 RT is active in transient transfection.";  
 RL J. Biol. Chem. 263:8384-8389(1988).  
 RN [4]  
 RP SEQUENCE OF 1391-1607 FROM N.A.  
 RX MEDLINE=85051302; PubMed=6209134;  
 RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;  
 RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of  
 RT coiled-coil alpha-helix.";  
 RL EMBO J. 3:2355-2362(1984).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.  
 RX MEDLINE=96196434; PubMed=8648630;  
 RA Stetefeld J., Mayer U., Timpl R., Huber R.;  
 RT "Crystal structure of three consecutive laminin-type epidermal growth  
 RT factor-like (LE) modules of laminin gamma1 chain harboring the

RT nidogen binding site.";  
 RL J. Mol. Biol. 257:644-657(1996).  
 RN [6]  
 RP STRUCTURE BY NMR OF 824-881.  
 RX MEDLINE=96196435; PubMed=8648631;  
 RA Baumgartner R., Caisch M., Mayer U., Poeschl E., Huber R.,  
 RA Timpl R., Holak T.A.;  
 RT "Structure of the nidogen binding LE module of the laminin gamma1  
 RT chain in solution.";  
 RL J. Mol. Biol. 257:658-668(1996).  
 CC -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ  
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF  
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING  
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
 CC -!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE  
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND  
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE  
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.  
 CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),  
 CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),  
 CC LAMININ-6 (K-LAMININ), AND LAMININ-7 (KS-LAMININ).  
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR  
 CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR  
 CC COMPONENT).  
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT  
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
 CC -!- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.  
 CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
 CC -!- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X05211; CAA28838.1; -;  
 DR EMBL; J03484; AAA39405.1; -;  
 DR EMBL; J02930; AAA39408.1; -;  
 DR EMBL; J03749; AAA39409.1; -;  
 DR PIR; A28469; MMSB2.  
 DR PDB; IKLO; 20-AUG-97.  
 DR PDB; ITLE; 12-FEB-97.  
 DR MGD; MGI:99914; Lamc1.  
 DR InterPro; IPR000122; Chemotaxis\_transducer.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001886; LamNT.  
 DR InterPro; IPR000034; Laminin\_B.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR Pfam; PF00052; laminin\_B; 1.  
 DR Pfam; PF00053; laminin\_EGF; 10.  
 DR Pfam; PF00055; laminin\_Nterm; 1.  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR PRODOM; PD002082; LamNT; 1.  
 DR PRODOM; PD003031; Laminin\_B; 1.  
 DR SMART; SM00180; EGF\_Lam; 9.  
 DR SMART; SM00001; EGF\_like; 1.  
 DR SMART; SM00281; LamB; 1.  
 DR SMART; SM00136; LamNT; 1.  
 DR PROSITE; PS00022; EGF\_1; 8.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 10.  
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.  
 FT SIGNAL 1 33  
 FT CHAIN 34 1607 LAMININ GAMMA-1 CHAIN.  
 FT DOMAIN 34 283 LAMININ N-TERMINAL (DOMAIN VI).  
 FT DOMAIN 284 502 4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN  
 FT V).  
 FT DOMAIN 284 339 LAMININ EGF-LIKE 1.

		544	544	D -> Y (IN REF. 2).
FT	CONFLICT	662	662	T -> S (IN REF. 2).
FT	CONFLICT	886	886	MISSING (IN REF. 2).
FT	CONFLICT	1158	1158	MISSING (IN REF. 2).
FT	CONFLICT	1434	1434	V -> A (IN REF. 2).
FT	CONFLICT	1475	1475	R -> K (IN REF. 4).
FT	CONFLICT	1576	1576	D -> N (IN REF. 4).
FT	CONFLICT	1576	1576	D -> N (IN REF. 4).
SQ	SEQUENCE	1607 AA;	177297 MW;	81B7B08E4869F242 CRC64;

  

Query Match            7.0%; Score 178.5; DB 1; Length 1607;  
 Best Local Similarity    21.5%; Pred. No. 2.1e-05;  
 Matches 103; Conservative 45; Mismatches 175; Indels 155; Gaps

Qy	15	INOIKSANCPVGTETNAGOVDDL-----GTPANCV-NCKKNFYNNAAAFVPGAS	64
Db	646	LNLNTSIRKIGTYSERTAYGLDDVTTLQSARPPGEVAFTVESCTCPVGGGQF----	CE 700
Qy	65	TCPCPOKKDAGAQPNPATANLIVTOCN---VKCPAGTAIAGGATDYAIIITEVCVNCRIN	121
Db	701	TCLPYRRETFPSLGYPSPC---VLCTCNHGHSETCDPETGVCCDEDNTAG--PHCEKCSDG	755
Qy	122	FYNPNAPFNAGASTCTACPVNRVGGALTAGNAATIIVAQCNI-----VACPQTGA-----	L 171
Db	756	YYGDSST---LGTSSDCQPCP-----CPGSSCAIIVPKTEVVCHTCPGTAGKKELC	805
Qy	172	DGVTVDYVSFECEVRCRLNFYNGNNGMTPEHPGKSQCTPCAIPAKPANVAOATLGNDA	231
Db	806	DDGFVGDPL-----GSNGFVRL-----CRPCQC--NDNIDPNAVGCN	841
Qy	232	TITAOQ-----NVA-----CPDGTISAAGVNNWVAON---TECTNCAPNFYN--NNAPNF	276
Db	842	RLTGECCLKCIYNTAFGYCDRCKEGFF-----GNLPADNPADKCKACACNPYGVVOQQSSC	896
Qy	277	NP--GNSCTLPCPANKD-----YGAETAAG-----GAATLAKOCNI-----ACPD	314
Db	897	NPVTGCCOCLPHVSGRDGCGTDPGIYYNLQSOGCERCDHALGSTNGQCIDRTGQCBQCP	956
Qy	315	GTAIASGATNYVIIOTELCNCAANFYFDGNNFOAGSRCKACPAKPVGAVATAGGTATL	374
Db	957	G-----ITGQHCCERC-----ETNHFGFEGEGCKPCDCHH-----EGSL	989
Qy	375	IQAQ-----ALECPACTVLTDTGTTSTYKQAASECYKCAANFYTKTDOWAGIDTCTSC	428
Db	990	SLOCKDDRCECREGFV-----GNRCQCEENFYFNRSW-PGQCECPAC	1032

  

<b>RESULT 12</b>			
<b>LMA1_HUMAN</b>			
ID	LMA1_HUMAN	STANDARD;	PRT; 3075 AA.
AC	P25391;		
DT	01-MAY-1992 (Rel. 22, Created)		
DT	01-MAY-1992 (Rel. 22, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	LAMININ ALPHA-1 CHAIN PRECURSOR {LAMININ A CHAIN}.		
GN	LAMA1 OR LAMA.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxId=9606;		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE+91333420; PubMed=1714537;		
RA	Haaparanta T., Uitto J., Ruoslahti E., Engvall E.;		
RT	"Molecular cloning of the cDNA encoding human laminin A chain.";		
RL	Matrix 11:151-160(1991). [2]		
RN	SEQUENCE OF I-2628 FROM N.A.		
RX	MEDLINE+91264789; PubMed=2049067;		
RA	Nissinen M., Vuolteenaho R., Boot-Handford R., Kallunki P.,		
RA	Tryggvason K.;		
RT	"Primary structure of the human laminin A chain. Limited expression		
RT	in human tissues";		

Biochem. J. 276:369-379(1991).

[3]

SEQUENCE OF 2397-3072 FROM N.A.  
MEDLINE=89280632; PubMed=2733383;  
Olson D., Nagayoshi T., Fazio M., Peltonen J., Jaakkola S.,  
Saaborn D., Sasaki T., Kulvanleim H., Chu M.L., Deutzmann R.,  
Timpl R., Uitto J.,  
Human laminin: cloning and sequence analysis of cDNAs encoding A, B1  
and B2 chains, and expression of the corresponding genes in human  
skin and cultured cells.,  
Lab. Invest. 60:772-782(1989).  
-1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ  
IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF  
CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING  
WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
-1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE  
DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND  
TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE  
COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.  
THE ALPHA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ) AND  
LAMININ-3 (S-LAMININ).  
-1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
-1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR  
COMPONENT).  
-1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT  
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
-1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.  
-1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
-1- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.  
-1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.  
-1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X58531; CAA1418.1; -  
PIR; S14458; S14458.  
HSP; P02468; 1TLE.  
MIM; 150320; -  
InterPro; IPR000561; EGF-like.  
InterPro; IPR001886; LamNT.  
InterPro; IPR000034; Laminin\_B.  
InterPro; IPR002049; Laminin\_EGF.  
InterPro; IPR001791; Laminin\_G.  
Pfam; PF00052; laminin\_B\_2.  
Pfam; PF00053; laminin\_EGF\_15.  
Pfam; PF00054; laminin\_G\_5.  
Pfam; PF00055; laminin\_Nterm; 1.  
PRINTS; PR00011; EGFLAMININ.  
ProDom; PD002082; LamNT; 1.  
ProDom; PD003031; Laminin\_B; 2.  
SMART; SM00180; EGF\_Lam; 14.  
SMART; SM00001; EGF\_like; 1.  
SMART; SM00281; LamB; 2.  
SMART; SM00282; LamG; 5.  
SMART; SM00136; LamNT; 1.  
PROSITE; PS00022; EGF\_1; 11.  
PROSITE; PS01186; EGF\_2; 2.  
PROSITE; PS01248; LAMININ\_TYPE\_EGF; 15.  
Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
SIGNAL 1 17 POTENTIAL.  
CHAIN 18 3075 LAMININ ALPHA-1 CHAIN.  
DOMAIN 18 269 LAMININ N-TERMINAL (DOMAIN VI).  
DOMAIN 270 516 4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN  
V).  
DOMAIN 270 326 LAMININ EGF-LIKE 1.  
DOMAIN 327 396 LAMININ EGF-LIKE 2.

FT	DISULFID	1016	1025	BY SIMILARITY.	LMA5 MOUSE	STANDARD;	PRT;	3635 AA.
FT	DISULFID	1028	1041	BY SIMILARITY.	AC	O61001;		
FT	DISULFID	1046	1056	BY SIMILARITY.	DT	01-NOV-1997 (Rel. 35, Created)		
FT	DISULFID	1046	1063	BY SIMILARITY.	DT	15-JUL-1998 (Rel. 36, Last sequence update)		
FT	DISULFID	1065	1074	BY SIMILARITY.	DT	15-JUL-1998 (Rel. 36, Last annotation update)		
FT	DISULFID	1077	1087	BY SIMILARITY.	DE	LAMININ ALPHA-5 CHAIN (FRAGMENT).		
FT	DISULFID	1403	1412	BY SIMILARITY.	GN	LAMA5.		
FT	DISULFID	1405	1419	BY SIMILARITY.	OS	Mus musculus (Mouse).		
FT	DISULFID	1422	1431	BY SIMILARITY.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
FT	DISULFID	1434	1449	BY SIMILARITY.	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
FT	DISULFID	1452	1466	BY SIMILARITY.	OX	NCBI_TaxID=10090;		
FT	DISULFID	1454	1476	BY SIMILARITY.	RN	[1]		
FT	DISULFID	1479	1488	BY SIMILARITY.	RP	SEQUENCE FROM N.A.		
FT	DISULFID	1491	1506	BY SIMILARITY.	RC	STRAIN=C57BL/6 X CBA; TISSUE=Lung;		
FT	DISULFID	1509	1521	BY SIMILARITY.	RX	MEDLINE=96081906; PubMed=7499364;		
FT	DISULFID	1511	1528	BY SIMILARITY.	RA	Miner J.H., Lewis R.M., Sanes J.R.;		
FT	DISULFID	1530	1539	BY SIMILARITY.	RT	*Molecular cloning of a novel laminin chain, alpha 5, and widespread		
FT	DISULFID	1542	1553	BY SIMILARITY.	RT	expression in adult mouse tissues.*;		
FT	DISULFID	1556	1560	INTERCHAIN (PROBABLE).	RL	J. Biol. Chem. 270:28523-28526(1995).		
FT	CARBOHYD	38	38	INTERCHAIN (PROBABLE).	RN	[2]		
FT	CARBOHYD	555	555	N-LINKED (GLCNAC. . .)	RP	REVISIONS.		
FT	CARBOHYD	665	665	N-LINKED (GLCNAC. . .)	RA	Miner J.H., Lewis R.M., Sanes J.R.;		
FT	CARBOHYD	763	763	N-LINKED (GLCNAC. . .)	RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.		
FT	CARBOHYD	926	926	N-LINKED (GLCNAC. . .)	CC	-1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ		
FT	CARBOHYD	952	952	N-LINKED (GLCNAC. . .)	CC	IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF		
FT	CARBOHYD	1045	1045	N-LINKED (GLCNAC. . .)	CC	CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING		
FT	CARBOHYD	1407	1407	N-LINKED (GLCNAC. . .)	CC	WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.		
FT	CARBOHYD	1479	1479	N-LINKED (GLCNAC. . .)	CC	-1- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF		
FT	CARBOHYD	1579	1579	N-LINKED (GLCNAC. . .)	CC	ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.		
FT	CARBOHYD	1596	1596	N-LINKED (GLCNAC. . .)	CC	-1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE		
CC					CC	TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE		
CC					CC	COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.		
CC					CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT		
CC					CC	MEMBRANES (MAJOR COMPONENT).		
CC					CC	-1- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND		
CC					CC	KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT		
CC					CC	AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.		
CC					CC	-1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT		
CC					CC	WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.		
CC					CC	-1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.		
CC					CC	-1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).		
CC					CC	-1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.		
CC					CC	-1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.		
CC					CC	-1- SIMILARITY: ALPHA-5 CHAIN IS MORE RELATED TO DROSOPHILA ALPHA		
CC					CC	CHAIN THAN TO MAMMALIAN ALPHA CHAINS 1-4.		
CC					CC	-----		
CC					CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC					CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC					CC	the European Bioinformatics Institute. There are no restrictions on its		
CC					CC	use by non-profit institutions as long as its content is in no way		
CC					CC	modified and this statement is not removed. Usage by and for commercial		
CC					CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC					CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC					CC	-----		
CC					CC	EMBL; U37501; AAC53430.1; .		
DR					DR	HSSP; P02468; I1LE.		
DR					DR	MGD; MGI:105382; Llama5.		
DR					DR	InterPro; IPR000361; EGF-like.		
DR					DR	InterPro; IPR001886; LamNT.		
DR					DR	InterPro; IPR000034; Laminin_B.		
DR					DR	InterPro; IPR002049; Laminin_EGF.		
DR					DR	InterPro; IPR001791; Laminin_G.		
DR					DR	Pfam; PF00052; laminin_B_1.		
DR					DR	Pfam; PF00053; laminin_Egf; 19.		
DR					DR	Pfam; PF00054; laminin_G; 2.		
DR					DR	Pfam; PF00055; laminin_Nterm; 1.		
DR					DR	ProDom; PD002082; LamNT; 1.		
DR					DR	ProDom; PD003031; Laminin_B; 1.		
DR					DR	SMART; SM00180; EGF_Lam; 17.		
DR					DR	SMART; SM00001; EGF_like; 2.		







FT	DOMAIN	1575	1775
FT	DOMAIN	1776	2111
FT	DOMAIN	1776	1808
FT	DOMAIN	1809	1858
FT	DOMAIN	1859	1916
FT	DOMAIN	1917	1969
FT	DOMAIN	1970	2016
FT	DOMAIN	2017	2063
FT	DOMAIN	2064	2111
FT	DOMAIN	2112	2697
FT	DOMAIN	2698	3712
FT	DOMAIN	2698	2862
FT	DOMAIN	2863	3048
FT	DOMAIN	3049	3223
FT	DOMAIN	3270	3296
FT	DOMAIN	3334	3528
FT	DOMAIN	3529	3712
FT	DOMAIN	2178	2249
FT	DOMAIN	2301	2321
FT	DOMAIN	2376	2450
FT	DOMAIN	2541	2676
FT	DISULFID	273	282
FT	DISULFID	275	296
FT	DISULFID	298	307
FT	DISULFID	310	330
FT	DISULFID	333	342
FT	DISULFID	335	367
FT	DISULFID	370	379
FT	DISULFID	382	400
FT	DISULFID	403	414
FT	DISULFID	405	421
FT	DISULFID	423	432
FT	DISULFID	435	445
FT	DISULFID	448	460
FT	DISULFID	450	468
FT	DISULFID	470	479
FT	DISULFID	482	492
FT	DISULFID	495	507
FT	DISULFID	497	514
FT	DISULFID	516	525
FT	DISULFID	528	538
FT	DISULFID	541	553
FT	DISULFID	543	560
FT	DISULFID	562	571
FT	DISULFID	574	584
FT	DISULFID	587	599
FT	DISULFID	589	605
FT	DISULFID	607	616
FT	DISULFID	619	629
FT	DISULFID	632	644
FT	DISULFID	634	650
FT	DISULFID	652	661
FT	DISULFID	664	674
FT	DISULFID	677	691
FT	DISULFID	679	700
FT	DISULFID	702	711
FT	DISULFID	714	729
FT	DISULFID	732	746
FT	DISULFID	734	753
FT	DISULFID	755	764
FT	DISULFID	767	782
FT	DISULFID	1375	1387
FT	DISULFID	1377	1394
FT	DISULFID	1396	1405
FT	DISULFID	1408	1418
FT	DISULFID	1421	1429
FT	DISULFID	1423	1436
FT	DISULFID	1438	1447
FT	DISULFID	1450	1453
FT	DISULFID	1466	1480
FT	DISULFID	1468	1487
FT	DISULFID	1489	1498

FT	DISULFID	1501	1511	BY SIMILARITY.
FT	DISULFID	1514	1526	BY SIMILARITY.
FT	DISULFID	1516	1533	BY SIMILARITY.
Query Match 6.8%; Score 173; Db 1; Length 3712;				
Best Local Similarity 20.9%; Pred. No. 0.0012;				
Matches 93; Conservative 47; Mismatches 161; Indels 144; Gaps				
QY	18	IKSANCPVGTETNAGQVDLDGTPANCVNCKNFYNNAAAFVPGASTCTCPQKKDAGA	77	
DB	1773	VERSCSPGGVSGH-----SCEDCAPGYHDPG---PYGGYCIPE-----	1809	
QY	78	QPNPPATANLVTCQN---VKPCAGTAIA---GGATDYAAIITECVNCRINFYNENAPN	129	
DB	1810	-----ECNGHSETCDCATGICSKQHGTEGDH-----CERCVSGYIG-NATN	1850	
QY	130	FNAGASTCTCPVNRVGGALTAGNAATIIVAQCNCVACPTGTALDDGVTYYVRSFT--	ECV 187	
DB	1851	GTPGDCMICACPL-----PFDSSNNFAT--SCEI-----SESGDQIHCECKPGVTPRCE	1897	
QY	188	KCLRNFYNGNGNWTNPENKSOCTPCPAIKPANVAQAATLGNDAITTAQCNCVACPDGTIS	247	
DB	1898	SCANGFY-----GEPEISIQ-VCKPCEC--SGNTINPEQSGCDTRTGEC-LRLNNTFG	1947	
QY	248	AAGVNNVVAQNTCTNCAPNFYNNAPNPGNSTCLPCPANKDYGAETAGGAATLAKQ	307	
DB	1948	AA-----CNLCAPGYGDAIKLN-----CQSCDCD-DLGTQ-----	1978	
QY	308	CNIACPDGTAIAGSATNYVILOTECLNCAANFYFDGNPFQAGSSRCKACPANKVOGAVAT	367	
DB	1979	--TCDFPVGVCTCHEN--VIGDRCDRCKPDHY---GFESGVG-CRACDC---GAASN	2024	
QY	368	AGGTATLIAQALECPAGTVLTDTGTSTTYKQAASCEVCKCAANFYTKTDWVAGIDTCT-	426	
DB	2025	STQCDPHTGHCA--CKSGV-----TCRQCDRCRAVD-----HMKYKDGCTP	2063	
QY	427	-SCNKKLTSGAEANLPESAKKNIQ	450	
DB	2064	CNCQGSYRFGGNC-PWTGK--CQC	2085	
RESULT	15			
YLK3_CABEL	ID	YLK3_CABEL	STANDARD;	PRT; 1895 AA.
AC	P4195L;			
DT	01-NOV-1995	(Rel. 32, Created)		
DT	01-NOV-1995	(Rel. 32, Last sequence update)		
DE	20-AUG-2001	(Rel. 40, Last annotation update)		
DE	PUTATIVE SERINE/THREONINE-PROTEIN KINASE D1044.3 IN CHROMOSOME III			
DE	(EC 2.7.1.-).			
GN	D1044.3.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	Pauley A., Waterston R.;			
RL	Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.			
CC	STRONG. TO ZC84.1.			
CC	-1- SIMILARITY: CONTAINS ? EGF-LIKE DOMAINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collabora			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstati			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no			
CC	modified and this statement is not removed. Usage by and for commu			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announ">http://www.isb-sib.ch/announ</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL: J000065; AAA50735.1. -			

```
DR HSP: Q63450; 1A06.
DR WormPep: D1044.3; CE01206.
DR InterPro: IPR002899; EB.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR Pfam: PF01683; EB; 12.
DR Pfam: PF00069; pkinase; 1.
DR SMART: SM00220; S_TKC; 1.
DR SMART: SM00289; WRI; 12.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS01186; EGF_2; 2.
KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
KW ATP-binding; EGF-like domain.
KW DOMAIN 431 703 PROTEIN_KINASE.
FT NP_BIND 437 445 ATP (BY SIMILARITY).
FT BINDING 461 461 ATP (BY SIMILARITY).
FT ACT_SITE 569 569 BY SIMILARITY.
FT ACT_SITE 569 569 BY SIMILARITY.
SQ SEQUENCE 1895 AA; 208393 MW; F23C9F7881353AD6 CRC64;

Query Match 6.7%; Score 171; DB 1; Length 1895;
Best Local Similarity 20.3%; Pred. No. 8.5e-05;
Matches 111; Conservative 53; Mismatches 220; Indels 162; Gaps 25;

QY 3 NNILVILII---SLFINQ--IKSANCVPGETNTAGQVDDLG--TPANCVNCOKNFYNN 55
DB 1190 NNQCILLSIVGETCIANQCCVGGAMCSGTCCQTNGATAMYGICISSSSCSNSNQVSIN 1249

QY 56 AAAF--VPGASTCTPCQKDGAGQNPAPTANLV--TQCNVKCPAG-TAAGGATDYAA 110
DB 1250 GMCYNTVQVGGSCSFSQ-----CLNNAVCINNICVSTFCVS CSTNQVCISNQCYNVS 1304

QY 111 IITEVCNCRINFYNENAPNFAGASTCTACPVNRVGGALTAGNAATIVAOQNVACPTGTA 170
DB 1305 IGSQCV-----GSOQCLS-----NSQCISSICQ--CPQGTQ 1333

QY 171 LDDGVTTDYRSFTCEVKRLNFYNGNGNGTFFNPGK-----SOCTPCPAIK-PANVAQA 225
DB 1334 QSNGVCT-----GNNNNNQOPNOVLINNOCYNTVSI GFQCFPQ 1375

QY 226 TLGNDATITAOQNVACPDGTISAAGV-----NNWVAQTECTNCAPNFYNNNAPN 275
DB 1376 CLGNSQCHNSMCQ--CPTGNTNVNGYCOGSGNGQNSQVLI NNQCYNTVSI GFQCFQ 1433

QY 276 FNPGNSTCL----PCPANKDYGAETAGGA-----ATLAKOCNIA-- 311
DB 1434 QCLGNSQCLNSICQCPSGSSNVNGYCOGSGNGQCNQNSNOVYNNQCYNTVPIGSQCQITQQ 1493

QY 312 -----CPDGT-----IASGATNVVILQTECLNCAANFYFDG 343
DB 1494 CLGNSQCMNSFCQCPSGPTNNVNFCTSSSSNLCSAGQTVOLDSSNOPINCLYSTCPNN 1553

QY 344 NNFOAGSSR-----CKACPANKVQGAATAGTATLIAQCALECPAGTVLTDGTTSTYKQ 398
DB 1554 SFCQYSSSQRYVCCRKCGTNS-SFQVSASQGVVTCFNS--QCASGYICSNGAC----- 1605

QY 399 AASECVKCAANFYTTKQTDWVAG-IDTCTSCNKKLTSGAEANLPESAKKNIQCD----- 451
DB 1606 ----CPNTNSNTCSTGTGTCFTGQISVGQCFCFNSVNI GDRQORSEQCLGGSCQNNLCQC 1661

QY 452 ----FAN 454
DB 1662 PNGFAN 1667
```